Top-down and bottom-up processes in speech comprehension.

Zekveld, A.A.; Heslenfeld, D.J.; Festen, J.M.; van Schoonhoven, R.

published in
NeuroImage
2006

DOI (link to publisher)
10.1016/j.neuroimage.2006.04.199

document version
Publisher's PDF, also known as Version of record

ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 32 (2006) 1826 – 1836

Top–down and bottom–up processes in speech comprehension ☆

Adriana A. Zekveld,[a] Dirk J. Heslenfeld,[a,*] Joost M. Festen,[b] and Ruurd Schoonhoven[c]

[a] *Department of Cognitive Psychology, VU University, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands*
[b] *Department of Audiology, VU University Medical Center, Amsterdam, The Netherlands*
[c] *Department of Physics and Medical Technology, VU University Medical Center, Amsterdam, The Netherlands*

Received 5 September 2005; revised 3 April 2006; accepted 19 April 2006
Available online 14 June 2006

Speech comprehension includes both bottom–up and top–down processes, and imaging studies have isolated a frontal–temporal network of brain areas active during speech perception. However, the precise role of the various areas in this network during normal speech comprehension is not yet fully understood. In the present fMRI study, the signal-to-noise ratio (SNR) of spoken sentences was varied in 144 steps, and speech intelligibility was measured independently in order to study in detail its effect on the activation of brain areas involved in speech perception. Relative to noise alone, intelligible speech in noise evoked spatially extended activation in left frontal, bilateral temporal, and medial occipital brain regions. Increasing SNR led to a sigmoid-shaped increase of activation in all areas of the frontal-temporal network. The onset of the activation with respect to SNR was similar in temporal and frontal regions, but frontal activation was found to be smaller than temporal activation at the highest SNRs. Finally, only Broca's area (BA44) showed activation to unintelligible speech presented at low SNRs. These findings demonstrate distinct roles of frontal and temporal areas in speech comprehension in that temporal regions subserve bottom–up processing of speech, whereas frontal areas are more involved in top–down supplementary mechanisms.

Keywords: Language processing; Speech intelligibility; Functional MRI; Fronto-temporal network; Brain

Several neuroimaging studies on the processing of auditory stimuli in the human brain have demonstrated a hierarchical organization of the temporal lobe. The processing of relatively simple stimuli (e.g., tones) is associated with activation in the primary auditory cortex (Heschl's gyrus [HG]) and the planum temporale (Binder et al., 1996, 2000), and the processing of more complex auditory stimuli such as speech is associated with activation in the superior temporal gyrus (STG) (Binder et al., 1994) and in particular the anterior superior temporal sulcus (STS) (Belin et al., 2000; Binder et al., 1994, 1996; Crinion et al., 2003; Price, 2000; Scott et al., 2000; Sharp et al., 2004; Specht and Reul, 2003). Increased complexity of speech stimuli is associated with increased activation in these areas, whereas increased complexity of nonspeech stimuli is associated with increased activation in the primary auditory cortex and anterior STG (Benson et al., 2001). In general, the literature demonstrates a dorsal-to-ventral hierarchical organization of auditory processing, with ventral and anterior areas in the STS involved in the processing of speech and dorsal and posterior regions involved in auditory processing regardless of stimulus complexity (Scott et al., 2000).

Some studies investigating hemodynamic responses to auditory stimuli have also examined frontal brain regions. In particular, the left inferior frontal gyrus (IFG), including Broca's area (Brodmann area [BA] 44 and BA45), seems to be involved in the processing of complex auditory stimuli. The data support a role for the left IFG in syntactic and phonological processing (Heim et al., 2003; Friederici et al., 2000). In addition, the area is associated with cognitive functions other than linguistic processing, such as working memory (Crinion et al., 2003; Müller et al., 2001; Specht et al., 2000).

The temporal and frontal areas involved in speech processing are highly connected; functional distinctions can be made between anterior and posterior routes connecting temporal and frontal brain regions (Scott and Johnsrude, 2003). The anterior route consists of the anterior belt and parabelt areas, the anterior STS, and ventrolateral and dorsolateral frontal cortex. In general, the anterior route seems predominantly involved in the processing of auditory patterns including speech, whereas the posterior route seems more involved in the processing of auditory spatial information (Rauschecker, 1998).

Specht and Reul (2003) have shown that intelligible words, when compared to sounds of animals and instruments, indeed activate an extended network of temporal and frontal regions. The intelligible speech used in that study differed from the other sounds on a number of parameters, so that activation in this network could be due to speech comprehension, but also to acoustic or phonetic speech-specific cues. To examine how speech comprehension affects activation in this frontal–temporal network, speech

☆ Adriana A. Zekveld is now at the Department of Audiology, VU University Medical Center, Amsterdam, The Netherlands. Ruurd Schoonhoven is now at Statistics Netherlands, The Hague, The Netherlands.

* Corresponding author. Fax: +31 20 888 8832.
E-mail address: dirk@psy.vu.nl (D.J. Heslenfeld).

Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.04.199

intelligibility should be varied while keeping acoustic differences between intelligible and unintelligible stimuli minimal. In a PET study, Scott et al. (2000) presented different types of speech stimuli with varying intelligibility. Intelligibility was manipulated by spectrally inverting (rotating) normal and "noise-vocoded" speech. Noise-vocoded speech is manipulated speech lacking phonetic features and intonation but is intelligible with some training. Rotated speech and rotated noise-vocoded speech are unintelligible. Activation in the anterior part of the STS was only observed during listening to intelligible speech, but the STG was activated regardless of the intelligibility of the speech (Scott et al., 2000). In an fMRI study extending these findings, Narain et al. (2003) observed that, in addition to the anterior STS, a region in the posterior temporal lobe also responded to intelligible speech. Sharp et al. (2004) also presented noise-vocoded speech in addition to normal speech. Activation evoked by making a semantic decision about the presented words was compared to activation while making a syllabic decision about the words. Semantic decision-making was shown to be associated with bilateral activation in the inferior temporal cortex (BA20), the temporal poles (BA38), and several frontal regions including BA9 and BA10.

In an fMRI study by Giraud et al. (2004), identical speech stimuli were presented that were unintelligible before training, but comprehensible with some effort after a training phase. Bilateral middle and inferior temporal regions were associated with speech comprehension. The left IFG (Broca's area) was associated with the effortful search for phonetic cues after training (once the stimuli were potentially meaningful). The anterior cingulate, the right middle frontal cortex (BA9), and both anterior insulae were related to stimulus complexity and were also activated by speech stimuli that required auditory search.

The aim of the present study is to test whether these areas are also activated during the comprehension of normal speech. One way to vary normal speech comprehension without changing the type of stimulus is to manipulate the intensity level or signal-to-noise ratio (SNR) of speech presented in noise. Using normal speech with varying SNRs while simultaneously measuring intelligibility can identify brain areas specifically involved in the comprehension of natural speech.

Binder et al. (2004) measured BOLD responses to spoken syllables masked with five levels of noise. One of two syllables was defined as a target syllable for each participant. Participants had to choose as fast as possible which one of two sequentially presented syllables was the target syllable. Sound identification led to activation in the auditory association cortex covering the anterolateral aspect of HG and adjacent STG. Surprisingly, they observed that some areas in the left IFG were increasingly active with decreasing SNR. The authors suggest that this may reflect the enhanced activation of internal representations of speech sounds (articulatory- or sound-based) as a template against which the stimuli were matched. It remains unclear whether a similar process takes place during normal speech comprehension.

Davis and Johnsrude (2003) aimed to distinguish areas in the frontal–temporal network involved in the bottom–up acoustic analysis of speech from areas involved in more higher-level, top–down linguistic processes. Several types of distorted speech were presented; each type was presented at three levels of intelligibility determined during a pilot study. Activation in bilateral temporal regions (auditory belt and parabelt areas) was correlated with intelligibility and dependent of the type of distortion. Activation in the region surrounding these areas (parabelt and more distant polymodal cortex) and activation in inferior frontal cortex were also correlated with intelligibility, but independent of the type of distortion. The authors suggest that the distortion-dependent areas are involved in the acoustic analysis of speech, whereas the distortion-independent areas are involved in speech processing at more abstract levels of representation. In addition, brain areas were identified in which activation increased as intelligibility decreased; a left lateralized frontal and temporal lobe system showed this profile. Davis and Johnsrude (2003) suggested that the distortion-independent frontal lobe areas that were increasingly activated by distorted speech are likely to be involved in nonacoustic compensatory mechanisms. In addition, they related increased activation in the distortion-dependent temporal lobe areas to the increased allocation of attention to speech input. Finally, another distortion-dependent area (frontal operculum) also showed increased activation with more degraded speech, but the study did not allow assessing the precise function of this region.

Although this study provides more insight into the specific role of areas in the frontal–temporal network involved in the bottom–up and top–down processing of speech, only three intelligibility levels were used for each type of distortion. Here, we present normal speech with a wider, more continuous range of intelligibility levels. In addition, unintelligible speech is included to enable a more direct association between brain activation and speech intelligibility. Colder and Tanenbaum (1999) also included speech in noise presented at relative low SNRs and tried to identify temporal lobe areas responding differently to detected words in noise compared to noise alone. However, word comprehension was not measured, some conditions included only very few trials, and the possibility of response biases or errors was not taken into account.

In summary, several studies have used speech stimuli masked with distinct levels of noise in order to manipulate speech intelligibility (e.g., Binder et al., 2004; Davis and Johnsrude, 2003; Colder and Tanenbaum, 1999; Scott et al., 2004). However, in these studies, only a limited range of SNRs were presented and comprehension was not always measured independently, so that detailed knowledge of the relationship between SNR, speech intelligibility, and activation of different regions in the frontal–temporal network is still limited.

The present study examined how the intelligibility of speech presented in noise affects the activation of brain areas involved in speech perception. To identify brain areas involved in speech perception, brain activation evoked by listening to sentences presented in noise was compared to activation evoked by listening to noise alone. The activation evoked by each sentence-in-noise stimulus was measured using an fMRI sparse-sampling paradigm. Speech intelligibility was examined by asking participants to identify (four-alternative forced choice) each stimulus directly after its presentation. In addition, directly following the scanning session, participants performed a surprise recognition task in order to determine their ability to recall the previously presented sentences. This task served as a control for the identification task performed during the scanning session. When participants erroneously misidentified a sentence during the scanning session (i.e., accidentally pressed the wrong button), they may recognize the sentence afterwards. Alternatively, correct guesses during identification will most likely not be remembered as heard sentences. We used a combination of identification and recognition in order to determine the intelligibility of each speech-in-noise stimulus (e.g., Plomp and Mimpen, 1979, Versfeld et al., 2000) in the MRI scanner.

Methods

Participants

Participants were 10 adults (4 males) aged 20–26 years (mean age was 22.7 years). All were right-handed on the Dutch 'Classificatie van links- en rechtshandige proefpersonen' (Classification of left- and right-handed subjects) (Van Strien, 1992) and were native speakers of Dutch. They did not report any hearing problems or a history of neurological disease. All participants provided written informed consent in accordance with the Ethical Committee of the VU University Medical Center, Amsterdam, The Netherlands.

Stimuli and tasks: scanning session

At the beginning of each trial, a red fixation cross appeared on the screen, indicating the start of the auditory stimulus 2000 ms later. The cross disappeared at the start of the binaurally presented auditory stimulus. This could be either a speech-in-noise or a noise-alone stimulus. Participants were unaware of the fact that noise-alone stimuli were presented; noise-alone stimuli could not be differentiated from speech-in-noise stimuli with the lowest SNRs, and participants had to perform the same task for each stimulus.

Speech-in-noise stimuli consisted of a spoken sentence masked with noise. The sentences were selected from a list of 507 everyday Dutch sentences of eight or nine syllables read by a female speaker (Versfeld et al., 2000).

For each participant, 48 of a total of 192 speech-in-noise trials were randomly selected during the experiment and replaced with noise-alone trials. The maximum number of successive noise-alone trials was two in order to prevent the task becoming too monotonous. The SNRs of the 144 remaining speech-in-noise trials ranged from −35 dB to 0 dB (4 participants) or from −30 dB to +5 dB (6 participants). These ranges covered unidentifiable to intelligible speech and resulted in the correct identification of approximately 50% of the presented sentences (which was determined during pilots in which the same identification task was presented). The noise level was fixed at 80 dB SPL; changing the intensity level of the spoken sentences varied SNR. This procedure ensured that the noise level at the beginning of a trial did not predict the SNR of the upcoming sentence. SNRs were created by dividing the SNR range (35 dB) by the number of speech-in-noise trials (144). The SNR of each speech-in-noise trial was randomly selected for each participant (without replacement). The sentence order was fixed across participants.

Noise with the long-term average spectrum of the speech served as mask (Versfeld et al., 2000). Noise onset was 500 ms before sentence onset, noise offset was 500 ms after the end of the sentence, and stimulus duration was 3000–4000 ms.

Noise-alone trials consisted of the same masking noise as used in the speech-in-noise trials. The duration of the noise was based on the duration of the speech-in-noise stimulus that was replaced by the noise-alone stimulus, resulting in noise-alone stimuli with the same length as speech-in-noise stimuli (3000–4000 ms).

Finally, all stimuli were prefiltered in order to correct for the transfer function of the auditory transducer headset (for details of the transducer see the MRI procedures section). The transfer function was measured with a 2 cm^3 coupler.

Functional activation was imaged 5500 ms after the beginning of each auditory stimulus. After the acquisition of the MRI volume (lasting 2000 ms), four response alternatives were visually presented on the screen. Participants responded by pressing the button corresponding to the sentence they had just heard. They were instructed to make a response on each trial. The four alternatives were numbered and presented on separate text lines; the right index finger corresponded to the uppermost response alternative, and the right little finger corresponded to the lowest alternative. The correct alternative (speech-in-noise trial) was randomly presented at one of the four possible locations. Participants had a maximum of 7000 ms to respond, and a time bar located at the bottom of the screen indicated the progression of the response time. After a response, or after the available response time, the response alternatives disappeared and the screen remained black until a fixation cross indicated the beginning of the next trial. The duration of this black screen consisted of the response time left over after the response, plus 3500 ms. Each trial lasted 20 s, and the 192 trials were presented in six blocks of 32 trials, each lasting 10.7 min. At the end of each block, participants were allowed a few minutes of rest.

This paradigm allowed stimulus presentation during quiet periods between scans and imaging of the peak of the BOLD response evoked by the auditory stimuli. The 20 s interval between scans prevented the imaged BOLD response from being confounded by the auditory stimulus and the scanner noise of the previous trial (sparse sampling paradigm) (Belin et al., 1999; Zatorre and Binder, 2000).

The four visually presented response alternatives consisted of the auditorily presented sentence (speech-in-noise trial) and three false alternatives. These alternatives were chosen from equivalent sets of sentences: subsets from Versfeld et al. (2000) and Plomp and Mimpen (1979). The selection of the false alternatives was based on the number of syllables, intonation, and tempo of the auditorily presented sentence, in such a manner that differences between alternatives were minimal. In addition, the contents of the sentences were taken into account; nouns and verbs in each alternative did not occur in any of the other sentences presented simultaneously. Thus, the false alternatives were similar to the auditorily presented sentence regarding the number of syllables, intonation, and tempo, but different regarding the words and contents of the sentence. This prevented participants from making a correct response based on information other than identified syllables or words, but once a (part of a) word was identified it was easy to choose the correct alternative. Thus, when an incorrect alternative was chosen, it was likely that the participant had been unable to identify any of the speech in the noise.

MRI procedures

Brain imaging was performed on a 1.5 T Siemens Sonata scanner (Siemens Medical Systems, Erlangen, Germany) equipped with a head volume coil. Functional volumes consisted of 20 near axial slices acquired using an EPI sequence with the following parameters: repetition time = 20 s, echo time = 60 ms, volume scan time = 2000 ms, flip angle = 90°, slice thickness = 5 mm, slice gap = 1 mm, acquisition matrix = 64 × 64 pixels, in-plane resolution = 3.3 × 3.3 mm. Images were on-line motion corrected.

After the functional imaging session, a three-dimensional structural scan was made using a T1-weighted MP-RAGE sequence with the following scanning parameters: repetition time = 2730 ms, echo time = 3.43 ms, inversion time = 1000 ms, flip angle = 7°, sagittal slice thickness = 1 mm, acquisition matrix = 224 × 256 pixels, in-plane resolution = 1 × 1 mm.

A notebook personal computer controlled stimulus presentation and response acquisition. Sound signals were generated with a Yamaha DS1 internal Sound Card (Yamaha, Hamamatsu, Japan). Stimuli were presented binaurally through an MRI-compatible STAX SR-001 MK2 electrostatic headset (STAX LTD., Saitama, Japan), connected with 15 cm plastic tubes coupled to the ear with foam earplugs for reduction of scanner noise. Visual stimuli were back-projected (Liesegang dv305, Düsseldorf, Germany) onto a screen that was viewed by the participants through an angled mirror positioned on top of the head coil. Participants responded by pressing fiber-optic buttons (Lumitouch Photon Control, Burnaby, Canada). Prior to the presentation of the first block of stimuli, participants were given a practice block to get familiar with the stimuli, response buttons, and their position in the scanner.

Stimuli and tasks: recognition task

Approximately 15 min after the end of the scanning session, participants performed a surprise recognition test in another, sound-proof room. The test was controlled by a personal computer. Signals were generated with a Creative SoundBlaster Sound Card (Creative Technology Ltd., Singapore, Singapore) and were presented binaurally through Sony MDR-V99 headphones (Sony, Tokyo, Japan).

Stimuli consisted of 288 sentences; 144 sentences were those presented auditorily during the scanning session (termed *old sentences*), the other 144 sentences were not previously presented. These *new sentences* were chosen from the same subset as the old sentences (Versfeld et al., 2000); none of them was used as response alternative in the identification task. The order of the sentences was random, the intensity level of the sentences was 70 dB SPL, and no masking noise was present.

After each stimulus, three labeled buttons located below each other appeared on the computer screen, and participants responded by clicking one of them with the mouse. The top button was labeled: "I haven't heard this sentence before" ("ik heb deze zin niet eerder gehoord"). This response will be referred to as a *no-response*. The second button was labeled: "I think that I have heard this sentence before" ("ik denk dat ik deze zin eerder gehoord heb"); this response will be referred to as a *familiar response*. The lowest button was labeled: "I know for sure that I have heard this sentence before" ("ik weet zeker dat ik deze zin eerder gehoord heb"); and this response will be referred to as a *recall response*. Participants were instructed to make a no-response when they experienced no feelings of familiarity during listening to the sentence. They were instructed to make a familiar response if they experienced some familiarity although were not able to recollect the sentence's occurrence during the scanning session. Finally, participants were instructed to make a recall response only if they were able to recollect the sentence's occurrence during scanning. After each response, the next stimulus was presented immediately. Duration of the recognition task was 30–45 min.

Behavioral analysis

Trials were coded depending on responses to speech-in-noise stimuli during the scanning session (identification task) and responses to old sentences during the recognition session. We defined intelligible sentences as sentences that were identified and recognized (*Id+Rec+*). Recognized sentences were the sentences to which a recall or a familiar response was made during the recognition task. We defined unintelligible sentences as sentences that were neither identified in the scanning task nor recognized afterwards (*Id−Rec−*). We constrained the category of intelligible sentences to those that were both identified and recognized in order to deal with the possibility of correct guesses during on-line identification. A frequently used measure of speech intelligibility is reproducibility (e.g., Plomp and Mimpen, 1979; Versfeld et al., 2000), which involves both identification and memory, and this combination of identification and recognition most closely resembles speech reproduction in a situation where speaking is undesirable.

To enable comparisons between specific subsets of intelligible and unintelligible sentences, a further subdivision of the Id+Rec+ and Id−Rec− categories was made based on the SNR of the sentences. We calculated the mean (over subjects) percentage of the sentences that were intelligible (Id+Rec+) for each SNR, sorted this percentage with increasing SNR, smoothed it with a moving average of 7 data points wide, and determined the SNR at which this percentage first exceeded chance level, for the SNRs above −30 dB (SNRs for which data of all participants were available). Chance level for identified and recognized sentences is 0.17 (chance level of the identification task times the chance level of the recognition task, i.e., 0.25 * 0.67). We refer to the SNR at which the percentage intelligible sentences first exceeded chance level as the chance-threshold. We then determined the SNR at which the percentage intelligible sentences first exceeded 70%, which we defined as the high-intelligibility threshold.

Sentences were divided into four categories. The first category contained unintelligible sentences that were presented at SNRs below the chance-threshold. These sentences will be referred to as *LOW−*. The second category contained unintelligible sentences presented at intermediate SNRs, that is, SNRs that exceeded the chance-threshold but were below the high-intelligibility threshold. These sentences will be referred to as *MID−*. The third category contained intelligible sentences at intermediate SNRs, that is, SNRs that exceeded the chance-threshold but were below the high-intelligibility threshold. These sentences will be referred to as *MID+*. The fourth category consisted of intelligible sentences that were presented at SNRs that exceeded the high-intelligible threshold. These sentences will be referred to as *HIGH+*.

A fifth category contained the noise-alone trials. The remaining sentences with ambiguous responses (i.e., sentences to which no response was made during the identification task, sentences that were identified and not recognized afterwards, and sentences that were not identified but recognized afterwards) were omitted from the analyses.

fMRI analysis

Preprocessing and statistical analysis of MRI data was performed using BrainVoyager 2000 software (Brain Innovation, Maastricht, The Netherlands). Preprocessing involved linear trend removal and three-dimensional spatial smoothing of each functional volume with a Gaussian kernel of 6 mm. Functional volumes were manually coregistered to the individual three-dimensional structural scans, resampled, and transformed into Talairach space (Talairach and Tournoux, 1988; Goebel et al., 2001).

Whole brain analysis

fMRI data were first analyzed for the whole brain and then specifically for a number of regions of interest (ROIs). The whole

Table 1
Mean, standard deviation, and range of the signal-to-noise ratios of the speech-in-noise stimuli in the four categories

	SNR		Range		
	M	SD	Minimum	Maximum	*N*
LOW−	−26.5	1.19	−35.0	−21.0	240
MID−	−14.2	0.58	−20.9	−3.1	246
MID+	−8.23	0.84	−20.8	−3.1	207
HIGH+	0.00	1.10	−2.9	5.0	189

N = total number of stimuli.

brain analysis served to identify brain regions that were more active during listening to intelligible speech in noise compared to listening to noise alone.

A multirun/multisubject design matrix was created specifying categories for each run and subject (e.g., Friston et al., 1999). Noise-alone trials were used as baseline in an initial fixed effects General Linear Model analysis. Constant term predictors (signal level confounds) were added for each run, and the voxel time courses of each run were *z*-transformed in order to stabilize their variances. The statistical threshold for the whole brain analysis was $P = 0.05$ (Bonferroni corrected).

ROI analyses

The whole brain analysis identified regions responding stronger to intelligible speech in noise than to noise alone. ROI analyses were then performed to reveal the details of the relationship between speech intelligibility and brain activity. ROIs were defined based on the whole-brain activation obtained in the HIGH+ category. Peaks of activation in each region were located, and significant voxels surrounding those peaks were selected and defined as ROIs. ROIs were limited in size (maximal 20 × 20 × 20 mm) in order to prevent voxels from being selected in multiple ROIs. Signal time courses were averaged and *z*-transformed for each ROI and run, and beta weights were obtained for each category and participant. These beta weights reflect the standardized strength of the neural response to stimuli in the various categories. For each ROI and category, the beta weights were then tested across participants with one-sample *t* tests (random effects analyses) to identify population responses to the speech-in-noise stimuli in the different categories. In addition, we directly compared the activation evoked by intelligible speech to the activation evoked by unintelligible speech for the mid-range of SNRs presented. That is, for each ROI, the beta weights of the MID+ category were compared to the beta weights of the MID− category with paired-sample *t* tests (random effects).

Results

Behavioral results

The SNR at which the mean percentage intelligible sentences first exceeded 16.67% (i.e., the chance-threshold) was −20.95 dB SNR. The SNR at which this percentage first exceeded 70% (i.e., the high-intelligible threshold) was −3.1 dB. Descriptive statistics for the four categories are given in Table 1.

We compared our behavioral data to those of Versfeld et al. (2000) who used the same stimuli but used reproduction as a measure of speech intelligibility. They found that 70% of the sentences can be completely reproduced when presented at −2.9 dB SNR. This is very close to our high-intelligibility threshold of −3.1 dB, supporting our classification of identified and recognized sentences as intelligible. We also fitted psychometric functions to the data of individual subjects and estimated the SNR at which 50% of the sentences were intelligible (i.e., were identified and recognized). The average of these estimated SNRs was −5.11 dB SNR. Versfeld et al. (2000) observed that the SNR at which 50% of these sentences can be correctly reproduced is −4.11 dB SNR, which again is close enough to support the current definition of intelligibility.

Participants were not encouraged to respond quickly, but they were instructed to make a response within the available response time. Reaction times were about 4500 ms for SNRs below −10 dB and then decreased almost linearly to about 1600 ms for SNRs of +5 dB.

Whole brain results

The whole brain analysis enabled us to identify brain regions responding to intelligible speech (statistical threshold $P = 0.05$, Bonferroni corrected). Compared to noise alone, intelligible speech in noise (i.e., HIGH+ stimuli) activated regions in the bilateral temporal cortex, the left frontal cortex, and medial occipital cortex. The coordinates of the activation peaks, the average *t* values, and the volumes of the activated regions are given in Table 2, and the areas are illustrated in Fig. 1.

Table 2
Significant ($P < 0.05$, Bonferroni corrected) activation foci for the HIGH+ stimuli obtained in the whole brain analysis, the Talairach and Tournoux (1988) coordinates of the activation peaks, the mean *t* values, and volumes of activated tissue

Region (BA)	Side	Peak coordinates			Mean *t* value	Volume (ml)
		x	*y*	*z*		
MFG (9)	L	−51	18	26	5.81	1.61
IFG pars triangularis (45)	L	−40	26	8	6.24	2.55
IFG pars opercularis (44)	L	−50	6	3	6.69	2.71
Anterior temporal (22, 41, 42)	L	−54	−14	5	10.13	7.90
Posterior temporal (22)	L	−54	−33	5	9.23	8.41
Anterior temporal (22, 42)	R	56	−7	1	9.88	7.78
Posterior temporal (22, 41, 42)	R	49	−28	7	8.19	7.73
Lingual gyrus (18, 19)	L	−11	−64	2	5.77	8.16
Lingual gyrus (18, 19)	R	11	−63	4	5.63	6.53

BA = Brodmann area; MFG = middle frontal gyrus; IFG = inferior frontal gyrus; L = left; R = right.

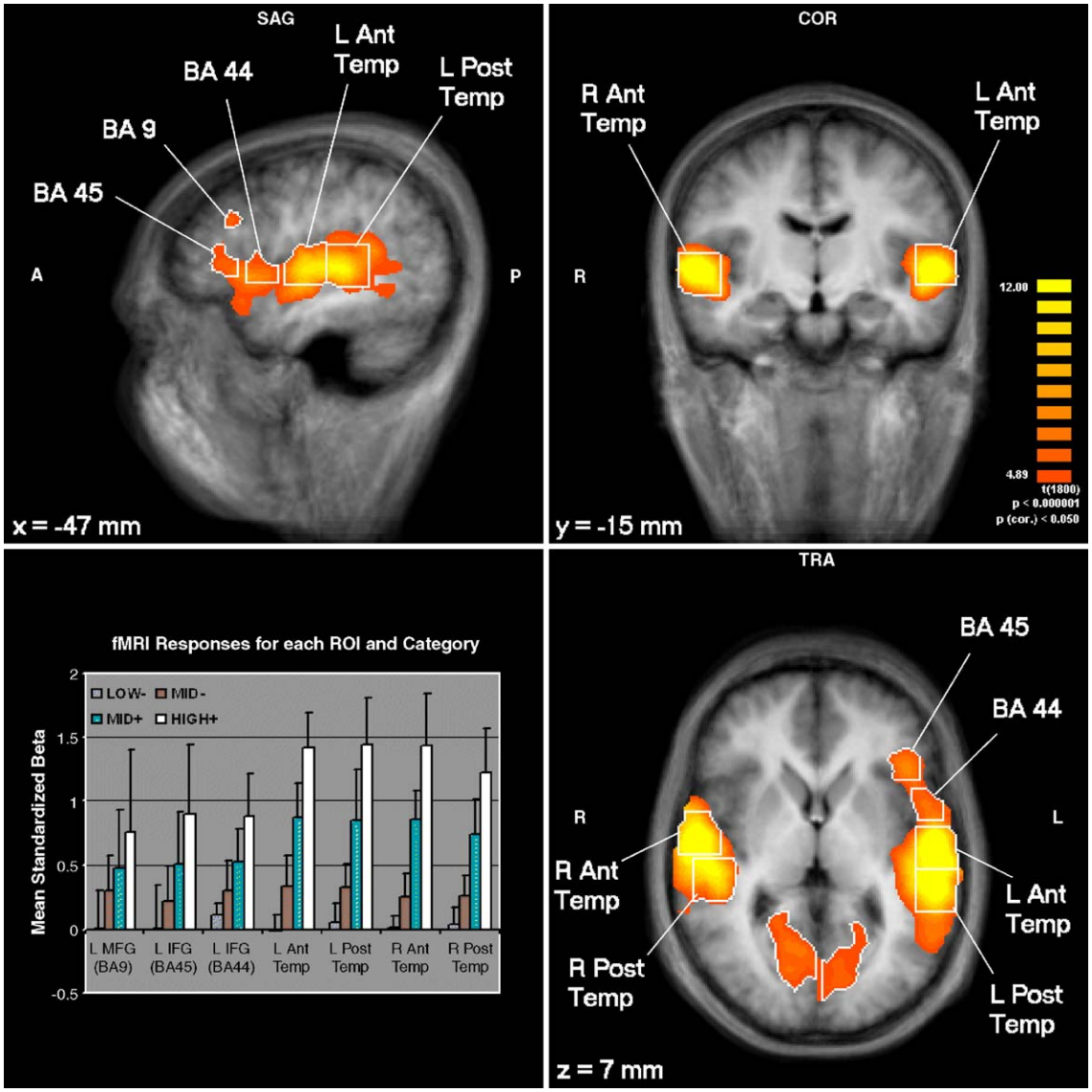

Fig. 1. Sagittal, coronal, and transversal maps showing activations and regions of interest obtained for intelligible speech stimuli (i.e., category HIGH+). The upper two images show the sagittal and coronal maps, the lower right image shows the transversal map. The views are at $x = -47$, $y = -15$, and $z = 7$ mm. The left of the brain is shown on the right of the images. The activation maps are superimposed on the averaged structural scan in Talairach space. The diagram in the lower left panel shows the mean and standard deviation of the beta weights obtained in all categories in the frontal and temporal ROIs. BA = Brodmann area; L = left; R = right; Ant = anterior; Post = posterior; Temp = temporal.

Frontal activations

Frontal activation consisted of three foci in the left hemisphere: pars opercularis (BA44) and pars triangularis (BA45) of the left IFG, and the left middle frontal cortex (mostly BA9). BA44 and BA45 are subregions of Broca's area.

Temporal activations

There were large areas of significant activation in both the left and right superior temporal lobe. The activation in the left hemisphere was larger and slightly more posterior than the activation in the right hemisphere. Significant activation was found in both anterior (BA22, BA38, BA41, and BA42) and posterior regions (BA22) of the left temporal lobe during listening to intelligible speech (i.e., HIGH+ stimuli). Significant activation evoked by intelligible speech was also observed in the anterior region of the right temporal lobe (BA22, BA38, and BA42).

Occipital activations

Intelligible speech evoked bilateral activation in the lingual gyrus (BA18/19).

ROI results

Based on the activation foci observed during listening to HIGH+ stimuli (see Table 2), nine ROIs were defined. The ROIs are illustrated in Fig. 1, together with the amount of activation obtained in all categories in the seven frontal and temporal ROIs (lower left panel).

For each participant and category, beta weights for the frontal and temporal ROIs were obtained and tested against zero across

Table 3
Results of the ROI analyses: mean, standard deviation (in parentheses), and significance of the standardized beta weights for each ROI and category

Region (BA)	Side	LOW−	MID−	MID+	HIGH+
MFG (9)	L		0.30**	0.48** (0.46)	0.76** (0.64)
IFG (45)	L		0.22* (0.27)	0.51** (0.41)	0.89** (0.56)
IFG (44)	L	0.11**(.10)	0.30** (0.23)	0.52** (0.26)	0.88** (0.33)
Ant temp (22, 41, 42)	L		0.33** (0.24)	0.87** (0.27)	1.41** (0.28)
Post temp (22)	L		0.32** (0.19)	0.84** (0.40)	1.45** (0.37)
Ant temp (22, 42)	R		0.26** (0.18)	0.85** (0.23)	1.43** (0.41)
Post temp (22, 41, 42)	R		0.26** (0.16)	0.74** (0.27)	1.22** (0.35)
Lingual gyrus (18, 19)	L			0.29* (0.32)	0.79** (0.55)
Lingual gyrus (18, 19)	R			0.32* (0.31)	0.76** (0.51)

Only results for significant beta weights are given.
Degrees of freedom of the one-sample *t* tests were 9. BA = Brodmann area; MFG = middle frontal gyrus; IFG = inferior frontal gyrus; L = left; R = right; Ant = anterior; Post = posterior; temp = temporal.
* $P < 0.05$.
** $P < 0.01$.

participants with one-sample *t* tests (random effects, $P < 0.05$) in order to determine the response within these regions to the stimuli in the four categories. The means, standard deviations, and the significance of the beta weights are given in Table 3. As was expected from the whole brain analysis, HIGH+ sentences led to highly significant activation also in these random-effects ROI analyses. Stimuli in the MID+ category also evoked significant activation in all ROIs and stimuli in the MID− category evoked activation in the seven frontal and temporal ROIs. Finally, unintelligible speech presented at low SNRs (i.e., the LOW− category) evoked significant activation only in the pars opercularis of the left IFG (BA44). The lingual gyri were not further analyzed because the activation observed in this region most likely reflects processes not directly associated with the intelligibility of the speech, such as visualization of the potential graphic structure of the words (cf. Démonet et al., 1994).

Finally, the activation evoked by intelligible speech in the MID+ category was directly compared to the activation evoked by unintelligible speech in the MID− category. Intelligible speech evoked significantly (random effects, $P < 0.05$) more activation in the temporal regions of both hemispheres. Furthermore, region BA45 of the left hemisphere was significantly more activated in response to intelligible speech as compared to unintelligible speech.

SNR analyses

Note that the standardized activation during listening to intelligible speech (both in the HIGH+ and MID+ categories) in the frontal ROIs is only about 60% of the standardized activation observed in temporal ROIs (i.e., for HIGH+, the mean betas are 1.38 for temporal and 0.84 for frontal regions; for MID+, they are 0.83 for temporal and 0.50 for frontal regions). The lower beta weights in frontal regions reflect the fact that these regions were relatively less activated during listening to intelligible stimuli. There are a number of possible causes for these lower beta weights in frontal regions, including: (*i*) frontal regions may start responding to intelligible sentences only at relatively high SNRs. Intelligible stimuli in the HIGH+ and MID+ categories covered a wide range of SNRs from −20.8 to +5 dB (see Table 1). The beta weights given in Table 3 are computed across all stimuli in each category. If frontal regions responded to speech stimuli only at relatively high SNRs, the resulting net beta weights would be smaller compared to the beta weights for the temporal regions. This implies a two-stage process in which temporal areas respond to detectable speech, but frontal areas only to intelligible speech at higher SNRs. This model is illustrated in the upper panel of Fig. 2. (*ii*) Alternatively, the frontal BOLD responses may be genuinely smaller for the entire range of SNRs from −20.8 to +5 dB. This would imply a one-stage process in which frontal regions are activated as soon as information is available in temporal cortex, but to a smaller degree. This model is illustrated in the lower part of Fig. 2.

To examine the origin of the smaller frontal activation, we ordered the standardized BOLD amplitudes by increasing SNR for each ROI and participant. Amplitudes for noise-alone trials were

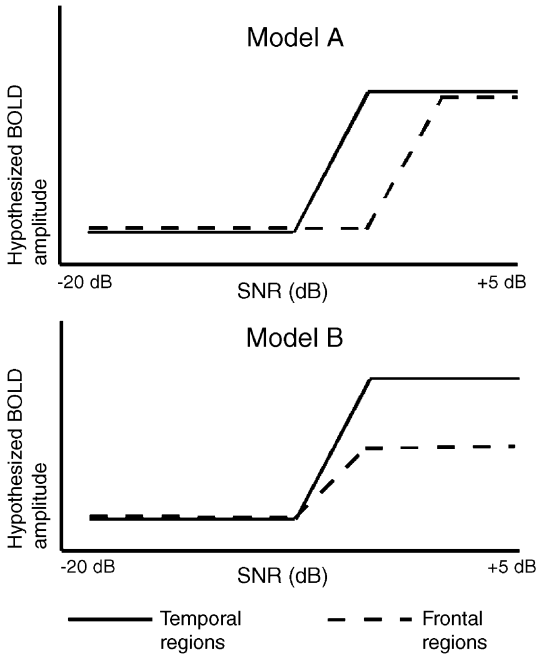

Fig. 2. Schematic illustration of models A and B. These models describe two possible causes for the lower mean beta weights evoked by intelligible speech in frontal compared to temporal regions. Model A (upper panel) implies a two-stage process with frontal regions responding only at relatively high signal-to-noise ratios (SNRs). Model B (lower panel) implies a one-stage process with genuinely smaller responses in frontal regions at higher SNRs.

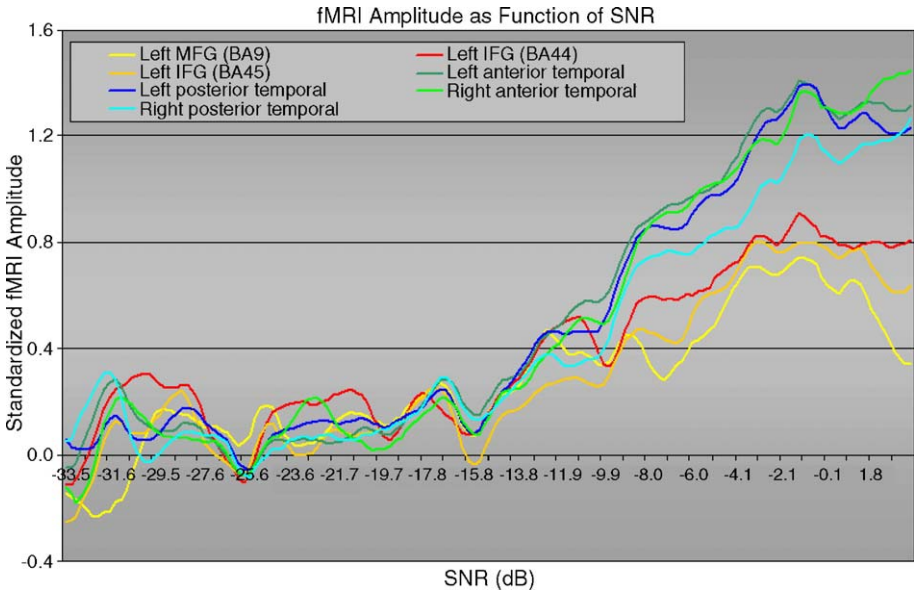

Fig. 3. Mean fMRI signal levels as a function of signal-to-noise ratio (SNR) separately for each region of interest (ROI). For each region, the standardized signal levels were ordered by increasing SNR for each participant. Amplitudes for noise-alone trials served as baseline and were subtracted from amplitudes for speech-in-noise trials. The resulting signal levels were averaged over participants for each ROI. BA = Brodmann area; MFG = middle frontal gyrus; IFG = inferior frontal gyrus.

treated as baseline and subtracted from amplitudes for speech-in-noise trials. The resulting developments of the standardized signal levels with SNR were averaged over participants for each ROI and are displayed in Fig. 3.

The difference in activation between the regions seems to be most pronounced at higher SNRs. The asymptotic amplitude of the activation is lowest in the frontal regions and highest in the temporal regions. Overall, the development of the BOLD response as a function of SNR seems sigmoid-shaped.

In order to test whether the onsets (model A) and/or the asymptotic response amplitudes (model B) of these functions differed, two additional analyses were performed. First, the onset of the activation was determined separately for each ROI and participant. The variability of the standardized amplitudes over the baseline (noise alone) trials was computed, and the onset was defined as the SNR at which the standardized amplitude first exceeded its 95% confidence interval. In Table 4 (left columns), means and standard deviations of the onset of activation in each ROI are given. There was no significant difference in the onset of activation between ROIs, $F(6,54) = 1.49$, Greenhouse–Geisser corrected $P = 0.25$, $\varepsilon = 0.34$.

Second, for each ROI and participant, standardized amplitudes were averaged over SNRs higher than the high-intelligibility threshold (−3.1 dB). These signal levels at the highest SNRs will be referred to as asymptotic amplitudes; their means and standard deviations over participants are given in Table 4 (right columns) for each ROI. Asymptotic amplitudes differed significantly between ROIs, $F(6,54) = 8.84$, Greenhouse–Geisser corrected $P < 0.001$, $\varepsilon = 0.42$.

Discussion

Consistent with the existing literature regarding auditory processing in the brain, the data of the present study show that the left frontal, bilateral temporal, and medial occipital cortices are involved in speech perception. Speech in noise evoked highly significant and spatially extended activation in these regions compared to noise alone. Intelligible speech resulted in more activation than unintelligible speech. Activation was observed in the middle frontal cortex (BA9) and in two parts of Broca's area: the pars opercularis (BA44) and pars triangularis (BA45) of the left

Table 4
Mean and standard deviation of the onset of the standardized fMRI activation function with increasing signal-to-noise ratio (SNR) (left columns), and the asymptotic standardized fMRI amplitude obtained to stimuli presented above the high-intelligibility threshold (>−3.1 dB) (right columns)

Region (BA)	Side	Onset of activation (dB SNR)		Asymptotic amplitude (Z score)	
		M	SD	*M*	SD
MFG (9)	L	−11.0	6.6	0.74	0.59
IFG pars triangularis (45)	L	−10.5	6.1	0.87	0.52
IFG pars opercularis (44)	L	−10.5	1.3	0.87	0.31
Ant temporal (22, 41, 42)	L	−13.6	1.7	1.4	0.28
Post temporal (22)	L	−13.5	2.1	1.4	0.33
Ant temporal (22, 42)	R	−13.1	1.9	1.4	0.36
Post temporal (22, 41, 42)	R	−13.4	2.1	1.2	0.28

BA = Brodmann area; MFG = middle frontal gyrus; IFG = inferior frontal gyrus; L = left; R = right; Ant = anterior; Post = posterior.

IFG. Activation observed in the left temporal lobe (anterior and posterior regions) was more extended and slightly more posterior than activation observed in the right temporal lobe. Based on the activation evoked by intelligible speech presented with relatively high SNRs, seven brain areas were identified to be involved in speech perception. These regions were then analyzed in detail to determine their role in speech comprehension. In these ROI analyses, the activation evoked by intelligible speech presented with intermediate SNRs was directly compared to the activation evoked by unintelligible speech presented with intermediate SNRs in order to examine which brain regions are more activated by intelligible speech compared to unintelligible speech when SNRs are comparable.

The first important finding of the current study is that only BA44 responded significantly to unintelligible speech at very low SNRs (SNRs less than −20.95 dB, LOW− category). It is well known that Broca's area is associated with both the production and perception of speech (Heim et al., 2003). In recent fMRI studies, decreasing SNRs of syllables and sentences presented in noise evoked increased activity in BA44 (Binder et al., 2004; Davis and Johnsrude, 2003). Binder et al. (2004) suggested that this finding reflects "enhanced activation of internal representations of the speech sounds (articulatory or sound based) as a template against which the sensory information can be matched" (p. 298). Our finding is consistent with a role of BA44 in the activation of internal speech representations. We did not examine detection of the presence of speech as such, but it is likely that participants were able to detect some speech cues in the noise during some trials in the LOW− category. This may have triggered a top–down activation of internal speech representations, which may serve to facilitate speech identification. Speech comprehension includes both bottom–up and top–down processes; task demands determine the relative weight and relevance of both processes (Rönnberg, 2003). When speech becomes less intelligible, listeners become more reliant on top–down working memory processes to infer what has been said (Pichora-Fuller et al., 1995). Broca's area is a likely candidate for such a top–down role in speech processing; several studies have supported its role in verbal working memory (Wilson, 2001). For example, Jäncke et al. (1998) found bilateral activation in the IFG in response to verbal and nonverbal stimuli. Consistent with the working memory literature (e.g., Baddeley, 1992), they suggested that this activation reflects the internal repetition of auditory information (i.e., articulatory loop of working memory). In addition, activation in Broca's area has been associated with the internal rehearsal of phonological codes in the absence of auditory stimulation (Paulesu et al., 1993). Combining this with the suggestions by Binder et al. (2004) leads us to conclude that activation in BA44 during listening to unidentified speech is likely to reflect a top–down process in which internal speech representations are activated and/or silently rehearsed. The fact that Broca's area is associated with speech production supports a role of BA44 in internal speech activation. Such top–down activation may serve to improve performance on an identification task.

This result is also in agreement with data showing that the left IFG is more activated in response to words compared to sounds (Specht and Reul, 2003). Specifically, the pars opercularis of the left IFG has been associated with the building of the syntactic structure of sentences (Newman et al., 2003). Moreover, BA44 is involved in phonological working memory, i.e., the activation and maintenance of phonological representations during both language production (Heim et al., 2003) and perception (Wagner et al., 2000). The involvement of BA44 in these processes is consistent with the observed activity in BA44 in response to intelligible speech, and it is plausible that these processes are already initiated prior to the identification of speech.

The second important finding of the present study is that several brain regions in the temporal and frontal cortex are significantly more activated by intelligible speech as compared to unintelligible speech at the same SNRs. Bilateral anterior and posterior superior temporal regions and area BA45 of the left IFG are significantly more activated by intelligible speech when this activation is directly compared to that evoked by unintelligible speech. This finding supports a role of these brain areas in the comprehension of speech in noise presented at SNRs between about −21 and −3 dB and is consistent with results of other studies examining speech comprehension (e.g., Crinion et al., 2003). In contrast to other studies (e.g., Sharp et al., 2004; Giraud et al., 2004), intelligible speech was not associated with activation in inferior temporal regions (BA21). This could be the result of signal loss due to fMRI susceptibility artifacts observed at tissue interfaces. The inferior temporal regions are prone to susceptibility artifacts (Devlin et al., 2000).

In the present study, the level of the speech was varied, and the level of the noise was kept fixed. This was necessary because otherwise the noise level would have predicted the SNR. However, it is unlikely that variations in the compound (signal + noise) intensity across the SNR range of the stimuli have influenced the results. The net variation of the compound intensity in the SNR range of the MID categories (−20.8 to −3.1 dB SNR) is only 1.7 dB SPL, which cannot account for the activation of the temporal and frontal brain regions by intelligible as compared to unintelligible speech. Also the net variation over the entire SNR range was very small, 80 to 86 dB SPL, about a mean of 83 dB SPL.

The third finding of the present study is that increasing SNR leads to a sigmoid-shaped increase of activation in both frontal and temporal cortices. The BOLD responses in these areas remain more or less constant until a certain SNR is reached. The onset of the BOLD responses as a function of SNR was highly similar between different brain regions, but their asymptotic amplitudes were different (i.e., our data support Model B, displayed at the bottom of Fig. 2). The absence of onset differences as function of SNR refutes a two-stage model in which frontal areas are activated only at higher SNRs. In contrast, our data suggest that speech intelligibility involves all areas in the network simultaneously. However, the differences in asymptotic amplitudes between areas may be related to differences in the degree of top–down and bottom–up processing. Smaller BOLD responses at high SNRs in frontal regions are in line with a top–down role in speech comprehension. At higher SNRs, participants identify a higher percentage of the sentences correctly. The lower reaction times at these higher SNRs also indicate that speech is more easily understood, so that there is less need for additional, auditory memory-related processes. In contrast, high amplitudes in temporal cortex may point to bottom–up processes, such as intensity coding of speech in noise. As can be seen in Fig. 3, the temporal BOLD responses as a function of SNR are similar for both hemispheres, indicating involvement of both temporal lobes in the processing of intelligible speech. This finding is consistent with other studies showing the involvement of bilateral, non-primary auditory temporal cortex in speech processing (Belin et al., 2000; Giraud et al., 2004). Based on the results of the SNR analyses, we

speculate that differences in the degree of top–down and bottom–up processing of speech may have caused the different asymptotic amplitudes of the BOLD responses in frontal and temporal regions.

Still, other SNR- or task-independent explanations for the differences in the magnitude of the activation observed in these brain regions must be considered: (*i*) the noise level of the raw MRI signals may be higher in frontal regions, thereby decreasing these standardized BOLD response estimates, and/or (*ii*) the latency or shape of the BOLD responses may differ between brain regions, leading to different response estimates in a sparse-sampling paradigm in which only one volume is recorded at a fixed latency (cf. Bellgowan et al., 2003; Handwerker et al., 2004; Henson et al., 2002). However, the data presented in Table 3 and Fig. 1 (lower left panel) argue against such aspecific or artifactual explanations because the frontal activation is not in all categories smaller than the temporal activation. In particular, the MID− category produced activation that was significant and of equal size in all frontal and temporal brain regions, thereby confining the smaller frontal response estimates specifically to intelligible speech (i.e., the MID+ and HIGH+ categories). Task-independent causes for the smaller frontal activation would have reduced the beta weights for all categories, not only those evoked by intelligible speech.

In conclusion, the present study demonstrates that the pars opercularis of the left IFG is activated when listening to unintelligible speech presented at low SNRs. This activation likely reflects the top–down generation and/or rehearsal of internal speech in order to facilitate the correct identification of the speech stimulus in the subsequent identification task. Consistent with Binder et al. (2004), it supports a role of BA44 in the activation of internal speech representations. In addition, bilateral temporal brain regions and BA45 in the left IFG are significantly more activated during listening to intelligible speech as compared to unintelligible speech. We furthermore show that that the development of the fMRI response as a function of SNR is sigmoid-shaped once speech intelligibility exceeds a certain threshold. The onset of the activation with increasing SNR is the same in different brain regions, but the peak amplitudes are lower in frontal compared to temporal cortex. These different peak amplitudes may reflect differences in the degree of top–down and bottom–up processing. The diminishing need for top–down processing when speech is clearly intelligible may have resulted in lower peak amplitudes in frontal regions, whereas ongoing bottom–up processes (such as intensity coding of the speech in noise) may have caused higher amplitudes of the BOLD response at high SNRs in temporal areas.

References

Baddeley, A., 1992. Working memory. Science 31 (255), 556–559.

Bellgowan, P.S.F., Saad, Z.S., Bandettini, P.A., 2003. Understanding neural system dynamics through task modulation and measurement of functional MRI amplitude, latency, and width. Proc. Natl. Acad. Sci. 100, 1415–1419.

Belin, P., Zatorre, R.J., Hoge, R., Evans, A.C., Pike, B., 1999. Event-related fMRI of the auditory cortex. NeuroImage 10, 417–429.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.

Benson, R.R., Whalen, D.H., Richardson, M., Swainson, B., Clark, V.P., Lai, S., Liberman, A.M., 2001. Parametrically dissociating speech and nonspeech perception in the brain using fMRI. Brain Lang. 78, 364–396.

Binder, J.R., Rao, S.M., Hammeke, T.A., Yetkin, F.Z., Jesmanowics, A., Bandettini, P.A., Wong, E.C., Estkowski, L.D., Goldstein, M.D., Haughton, V.M., Hyde, J.S., 1994. Functional magnetic resonance imaging of human auditory cortex. Ann. Neurol. 35, 662–672.

Binder, J.R., Frost, J.A., Hammeke, T.A., Rao, S.M., Cox, R.W., 1996. Function of the left planum temporale in auditory and linguistic processing. Brain 119, 1239–1247.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512–528.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Douglas Ward, B., 2004. Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. 7, 295–301.

Colder, B.W., Tanenbaum, L., 1999. Dissociation of fMRI activation and awareness in auditory perception task. Brain Res. Cogn. Brain Res. 8, 177–184.

Crinion, J.T., Lambon-Ralph, M.A., Warburton, E.A., Howard, D., Wise, R.S.J., 2003. Temporal lobe regions engaged during normal speech comprehension. Brain 126, 1193–1201.

Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. J. Neurosci. 23, 3423–3431.

Démonet, J.-F., Price, C., Wise, R., Frackowiak, R.S.J., 1994. A PET study of cognitive strategies in normal subjects during language tasks. Brain 117, 671–682.

Devlin, J.T., Russell, R.P., Davis, M.H., Price, C.J., Wilson, J., Moss, H.E., Matthews, P.M., Tyler, L.K., 2000. Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. NeuroImage 11, 589–600.

Friederici, A.D., Meyer, M., Von Cramon, D.Y., 2000. Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. Brain Lang. 74, 289–300.

Friston, K.J., Holmes, A.P., Price, C.J., Buchel, C., Worsley, K.J., 1999. Multisubject fMRI studies and conjunction analyses. NeuroImage 10, 385–396.

Giraud, A.L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M.O., Preibisch, C., Kleinschmidt, A., 2004. Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cereb. Cortex 14, 247–255.

Goebel, R., Muckli, L., Zanella, F.E., Singer, W., Stoerig, P., 2001. Sustained extrastriate cortical activation without visual awareness revealed by fMRI studies of hemianopic patients. Vision Res. 41, 1459–1474.

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21, 1639–1651.

Henson, R.N.A., Price, C.J., Rugg, M.D., Turner, R., Friston, K.J., 2002. Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. NeuroImage 15, 83–97.

Heim, St., Opitz, B., Müller, K., Friederici, A.D., 2003. Phonological processing during language production: fMRI evidence for a shared production–comprehension network. Brain Res. Cogn. Brain Res. 16, 285–296.

Jäncke, L., Shah, N.J., Posse, S., Grosse-Ryuken, M., Müller-Gärtner, H.-W., 1998. Intensity coding of auditory stimuli: an fMRI study. Neuropsychologia 36, 875–883.

Müller, R.-A., Kleinhans, N., Courchesne, E., 2001. Broca's area and the discrimination of frequency transitions: a functional MRI study. Brain Lang. 76, 70–76.

Narain, C., Scott, S.K., Wise, S.T., Rosen, S., Leff, A., Iversen, S.D., Matthews, P.M., 2003. Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb. Cortex 13, 1362–1368.

Newman, S.D., Just, M.A., Keller, T.A., Roth, J., Carpenter, P.A., 2003. Differential effects of syntactic and semantic processing on the subregions of Broca's area. Brain Res. Cogn. Brain Res. 16, 297–307.

Paulesu, E., Frith, C.D., Frackowiak, R.S.J., 1993. The neural correlates of the verbal component of working memory. Nature 362, 342–345.

Pichora-Fuller, M.K., Schneider, B.A., Daneman, M., 1995. How young and old adults listen to and remember speech in noise. J. Acoust. Soc. Am. 97, 593–608.

Plomp, R., Mimpen, A.M., 1979. Improving the reliability of testing the speech reception threshold for sentences. Audiology 18, 43–52.

Price, C.J., 2000. The anatomy of language: contributions from functional neuroimaging. J. Anat. 197, 335–359.

Rauschecker, J.P., 1998. Cortical processing of complex sounds. Curr. Opin. Neurobiol. 8, 516–521.

Rönnberg, J., 2003. Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. Int. J. Audiol. 42, S68–S76.

Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26, 100–107.

Scott, S.K., Blank, C.C., Rosen, S., Wise, J.S., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400–2406.

Scott, S.K., Rosen, S., Wickham, L., Wise, R.J.S., 2004. A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. J. Acoust. Soc. Am. 115, 813–821.

Sharp, D.J., Scott, S.K., Wise, R.S.J., 2004. Retrieving meaning after temporal lobe infarction: the role of the basal language area. Ann. Neurol. 56, 836–846.

Specht, K., Reul, J., 2003. Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task. NeuroImage 20, 1944–1954.

Specht, K., Shah, N.J., Jäncke, L., 2000. Bilateral inferior frontal networks are involved in speech perception processes. NeuroImage 11, S292.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme, New York.

Van Strien, J.W., 1992. Classificatie van links- en rechtshandige proefpersonen. Ned. Tijdschr. Psychol. 47, 21.

Versfeld, N.J., Daalder, L., Festen, J.M., Houtgast, T., 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J. Acoust. Soc. Am. 107, 1671–1684.

Wagner, A.D., Koutstaal, W., Maril, A., Schacter, D.L., Buckner, R.L., 2000. Task-specific repetition priming in left inferior prefrontal cortex. Cereb. Cortex 10, 1176–1184.

Wilson, M., 2001. The case for sensorimotor coding in working memory. Psychon. Bull. Rev. 8, 44–57.

Zatorre, R.J., Binder, R.J., 2000. Functional and structural imaging of the human auditory system. In: Toga, A.W., Mazziotta, J.C. (Eds.), Brain Mapping the Systems. Academic Press, San Diego, CA, pp. 365–402.